Design Concepts for Resilient Database Replication In Tenuous Communication Environments

NATO IST TG-12 Workshop:

Data Replication Over Disadvantaged Tactical Communication Links

11 September 2002

Sam Chamberlain, Ph.D.
US Army Research Laboratory
Computational & Information Sciences Directorate
Computer & Communication Sciences Division
wildman@arl.army.mil
(410) 278-8948 // DSN 298
http://www.arl.army.mil/~wildman

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **01 DEC 2007**	2. REPORT TYPE **N/A**	3. DATES COVERED
4. TITLE AND SUBTITLE **Design Concepts for Resilient Database Replication In Tenuous Communication Environments**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **US Army Research Laboratory Computational & Information Sciences Directorate Computer & Communication Sciences Division**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR’S ACRONYM(S) 11. SPONSOR/MONITOR’S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release, distribution unlimited.**		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**31**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

5 Reason Not to Worry About This Problem

5. There is too much data already –
   soldiers can't handle it.
4. Bandwidth is not a problem
   (It's just a modulation problem).
3. High speed radios will soon permeate the battle field.
2. Nothing will ever replace USMTF.
1. Industry will fix this problem.

More Why Not Worry About This ….

From: ***Computer Networks*** **(3rd ed., 1996), by Andrew S. Tanenbaum:**

"In the race between computing and communication, communication won. The full implications of essentially infinite bandwidth (although not at zero cost) have not yet sunk into a generation of scientists and engineers taught to think in terms of low Nyquist and Shannon limits imposed by copper wire. The new conventional wisdom should be that all computers are hopelessly slow, and networks should try to avoid computation at all costs, no matter how much bandwidth that wastes. In this section, we shall study fiber optics …"

This statement may be true for fiber optics, but it is often repeated as a general rule. However, wireless communications is bounded by the above limits and, at the lowest echelons, is often a limiting factor in effective battle command.

In other words, we always want to send more than we can.

Commercial Information Management Is Based Upon Guided (e.g., Wire & Fiber) Communications

Relative Performance	Computer Technology (Single CPU)	Communications Technology
1970's:	*100 nsec / instr. (CDC 6600)*	*56 Kbps (Arpanet)*
1990's:	*1 nsec / instr (Cray)*	*1 Gbps (Fiber)*
2000's:	*.1 nsec / instr (Multi)*	*50 Gbps (Fiber)*
Improvement:	*10 fold / decade*	*50-100 fold / decade*

… and 50,000 Gbps (50 Tbps) is attainable with current fiber technology.

Bit Error Rates:

99.64% of bit errors over fiber (telephone system) are single bit errors[1] and can be handled by a simple checksum. With an ATM, 8-bit, header checksum, the probability of not detecting a bad cell header is 10^{-20} or one cell every 90,000 years at OC-3 rates (155.52 Mbps).
(Note: w/ 1 billion ATM telephones, this is 1000 cell errors / year).

Note: This is NOT a Tactical Internet Environment.

1. *Observations of Error Characteristics of Fiber Optic Transmission Systems*; CCITT SG XVIII, San Diego, Jan 89.

In the Mobile Wireless Environments Computational Power Is Outpacing Communications Power

EXAMINE THE RATIO OF: $\dfrac{\textit{COMPUTING POWER (MFLOPS)}}{\textit{COMMUNICATIONS POWER (Mbps)}}$

WIRED (e.g., LAN) vs. WIRELESS (e.g., radio) ENVIRONMENTS

COMPUTER	(MFLOPS)*	COMM SYS.	(Mbps)	RATIO (MFLOPS/Mbps)	
Pentium 4/2000	(655)	Gbit Ethernet	(1000)	.65	GUIDED
Pentium III/550	(197)	Ethernet	(100)	1.97	
Pentium 233 MMX	(33)	Ethernet	(10)	3.3	
Pentium 4/2000	(655)	JTRS	(20)	32.7	WIRELESS
"		NTDR	(.375)	1,746	
"		SINCGARS	(.0064)	5,156	

* The Vector Whetstone-97 Benchmark. http://www.dl.ac.uk/TCSC/disco/Benchmarks/whetstone.html

CONCLUSION: FOR BATTLE COMMAND AT THE FIGHTING ECHELONS, WHERE WIRELESS COMMUNICATIONS IS THE NORM, WARRIORS NEED TO COMPLEMENT COMMUNICATIONS POWER WITH CAREFUL INFORMATION MANAGEMENT.

Observations and Basic Points

- **Just as important as the actual bandwidth values are the huge variation of the bandwidth.**
- **These parameters have significant implications for building battle command systems – especially at the lowest echelons.**
- **In the battle command business, one must consider the whole environment when addressing the problems associated with data replication.**
- **0 is a valid value for throughput [aka delay = ∞].**
- **Interesting requirement:**
  One must be able to operate when throughput = 0,
  Otherwise, just throw away your battle command system.
- **Usual response: Are you nuts?**

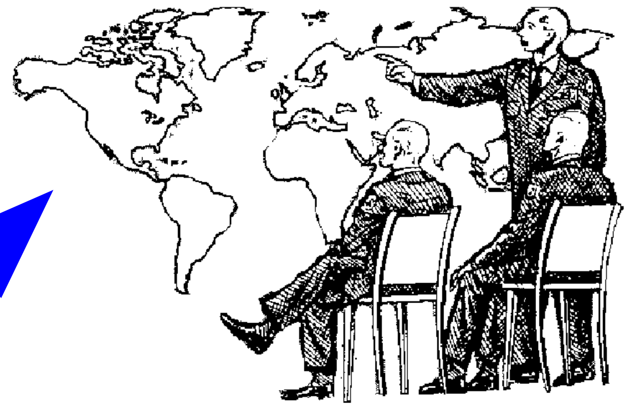
Some Assumptions and Tenets

- **Computers will get faster, smaller and use less power.**
- **Tactical communications will continue to be**
  - **Highly variable (Guided & Wireless; 0 bps to Gbps), and**
  - **Often less than demanded**
- **There will be more information to send than can be sent; we have to quickly select what and when we exchange.**
- **Warriors under fire don't have time to fuss with computers.**
- **We can't *afford* to propagate message-based, legacy systems (like TF-XXI). We need Common Distributed Computing Environments with Model-Based Battle Command.**

Therefore: Information Management & Distribution Must Be:

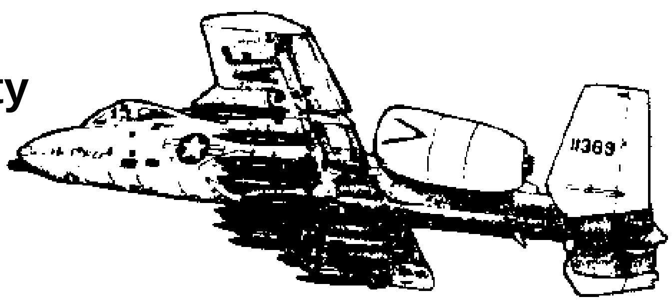
- **Automatic (Hands-Off, Context-Based)**
- **Adaptive (Context-Based, Respond to Bandwidth Conditions)**
- **Affordable (Treated as a Distributed Computing Environment)**

Affordability: Ultimate Goal - Common Environments At All Echelons

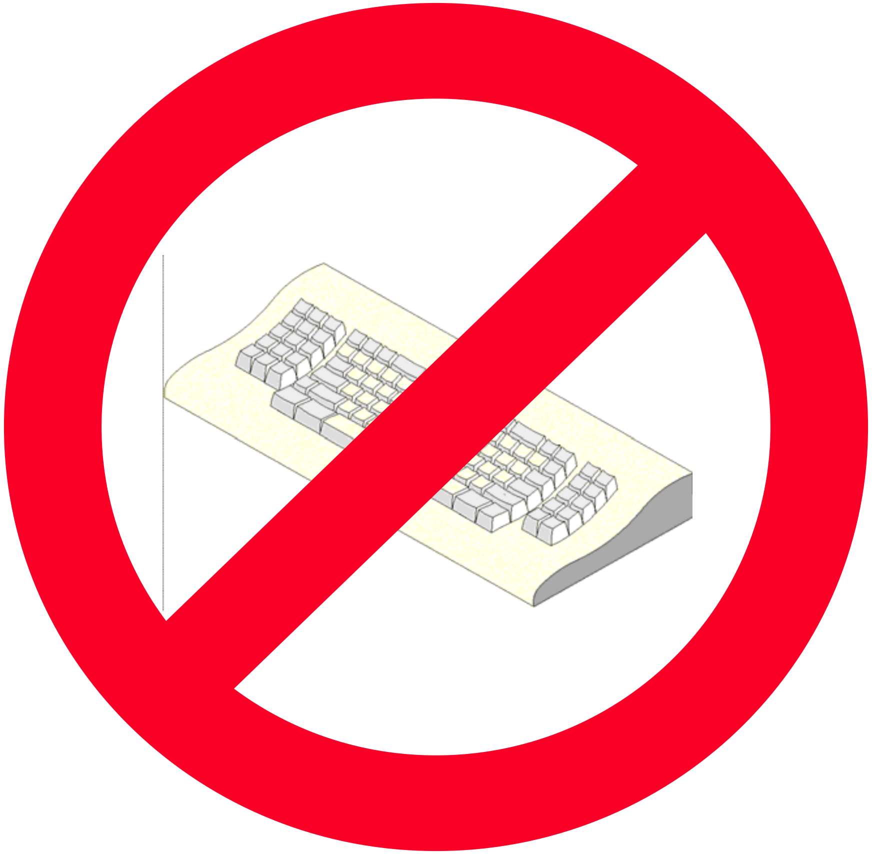


An Environment With True "CINC-to-Foxhole"

Capability and Applicability



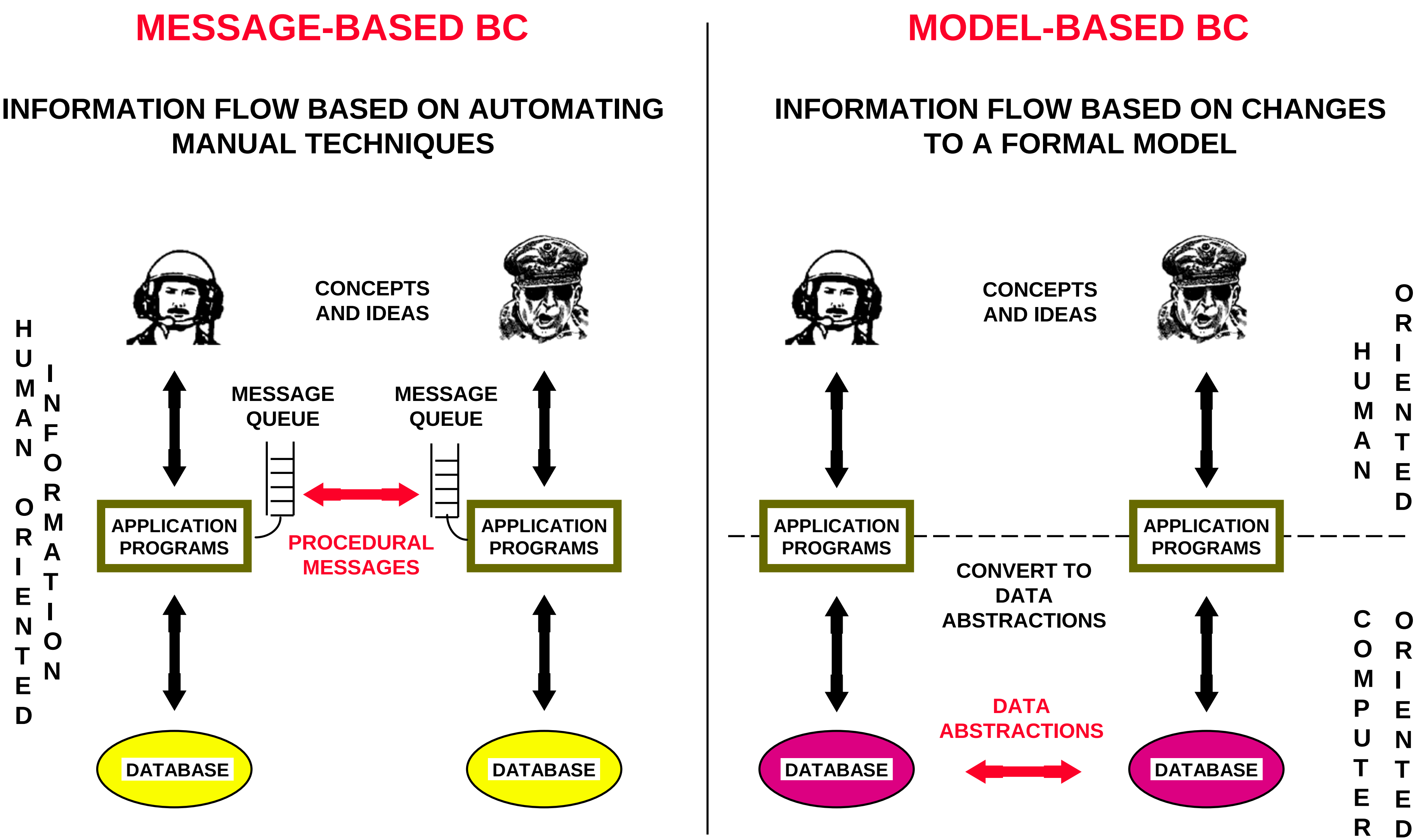
MAJOR PITFALL OF EVERY ATTEMPT TO AUTOMATE BATTLE COMMAND TO DATE



THE AUTOMATION OF MANUAL TECHNIQUES

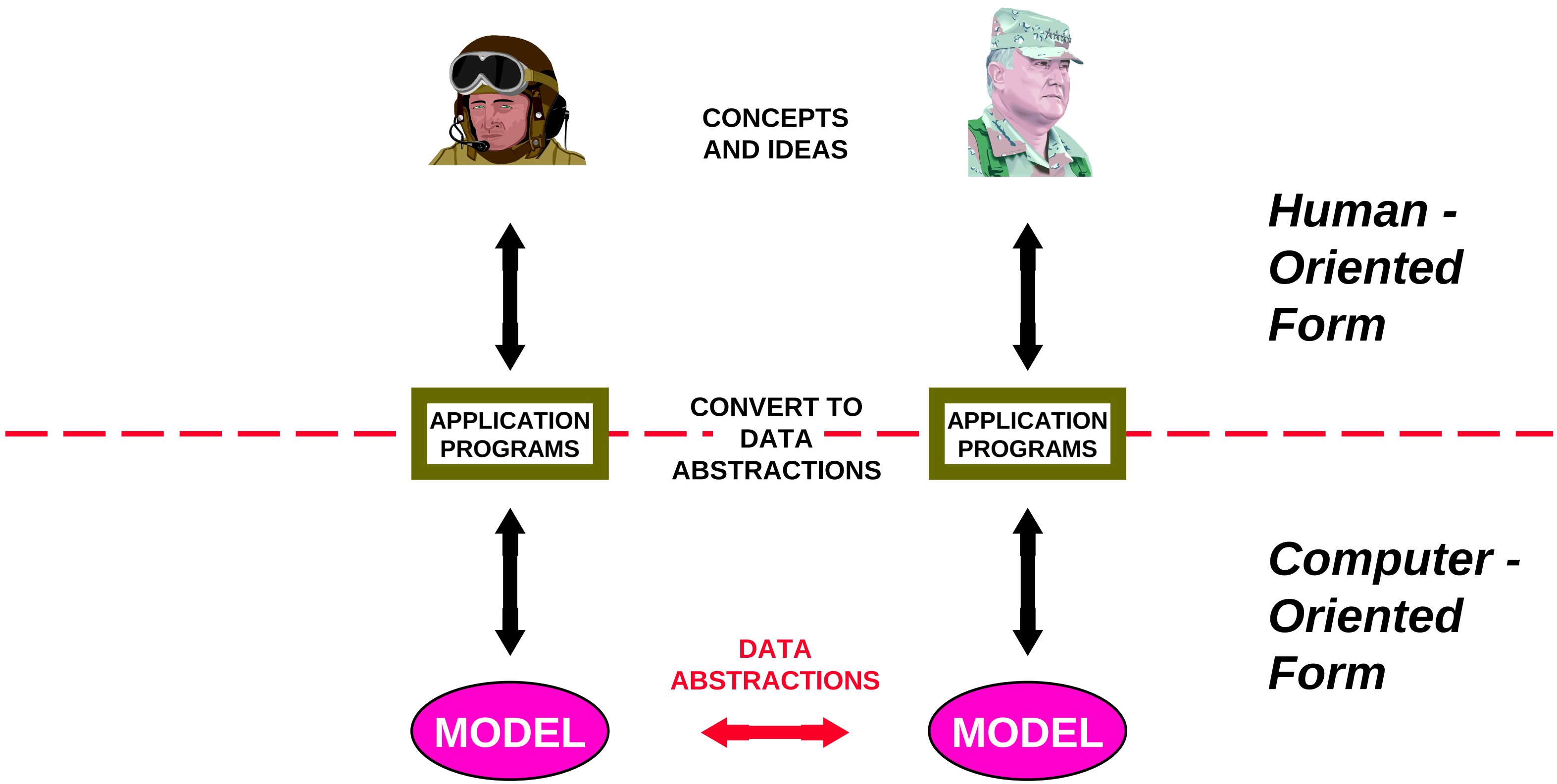
Need to rid ourselves of the “E-Mail” mentality

Model vs. Message-Based Battle Command (BC): This Is What Makes Automation and Adaptation Affordable



UNCLASSIFIED – APPROVED FOR PUBLIC RELEASE

Model-Based Battle Command: Each Node Maintains A Model Of Its Perception Of The Battlefield



*** Data Abstractions Are The Medium Of Communications ***

*** Exchange Controlled by Active Database Triggers (State of Database) ***

*** Reasonably Different Perceptions Of The Battlefield Are Allowed ***

*** Synchronization Is The Realistic Control Of Differing Perceptions ***

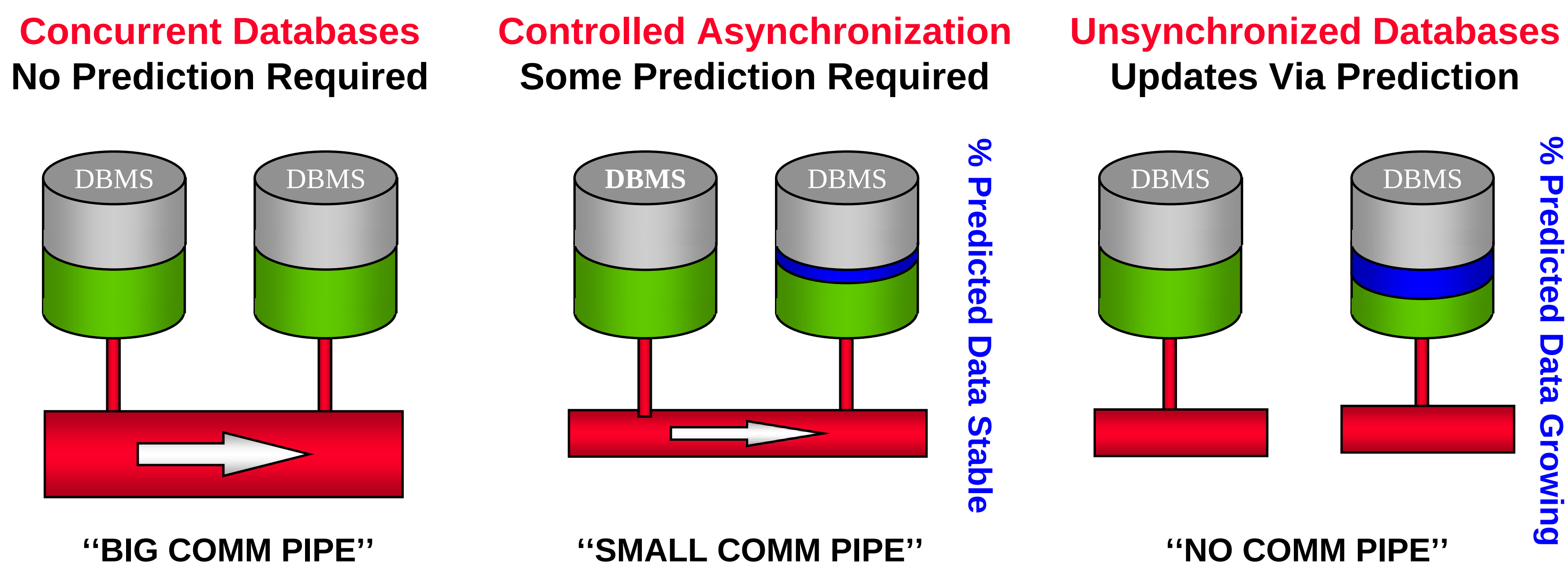
UNCLASSIFIED – APPROVED FOR PUBLIC RELEASE

Some Ideas Up To This Point

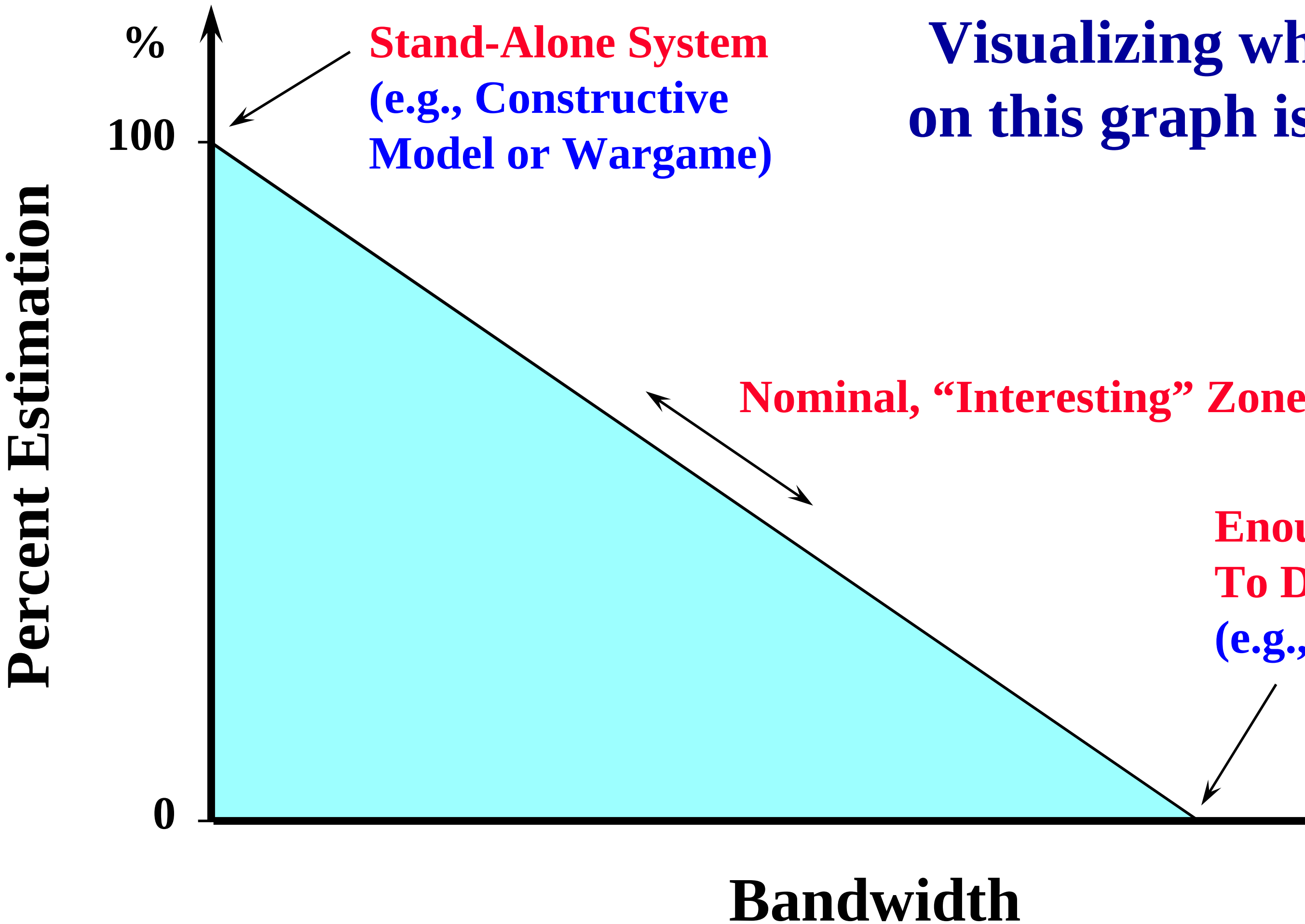
- *Synchronization* Is The Realistic Control Of Differing Perceptions.
- What Do We Want? . . . Perfect Synchronization, Of Course!
- Are We Going To Get It? . . . Not Likely Most Of The Time.
- Ok, Look At Extremes: + Infinite Bandwidth - Perfect Synchronization.
  - No Bandwidth - Unknown Synchronization.
- In The Commercial World, We Tend To Build Systems Toward Perfect Synchronization, But This Is Not Realistic For Most Military Environments.
- In The Military Environment We Must Be Able To Handle Both Extremes, Infinite Bandwidth And No Bandwidth, Plus All The Interesting Cases In Between; This Is A Form Of Adaptive Battle Command.
- Therefore, We Must:
  - Mix Predictive Modeling with Known Synchronized Information and Keep the User Aware of Which Is Which;
  - Monitor (Or Measure) Bandwidth and Control Synchronization Based on Current Communication Resources.

Adaptive Synchronization Based On Bandwidth

0 is a valid value for throughput, or ∞ for delay!

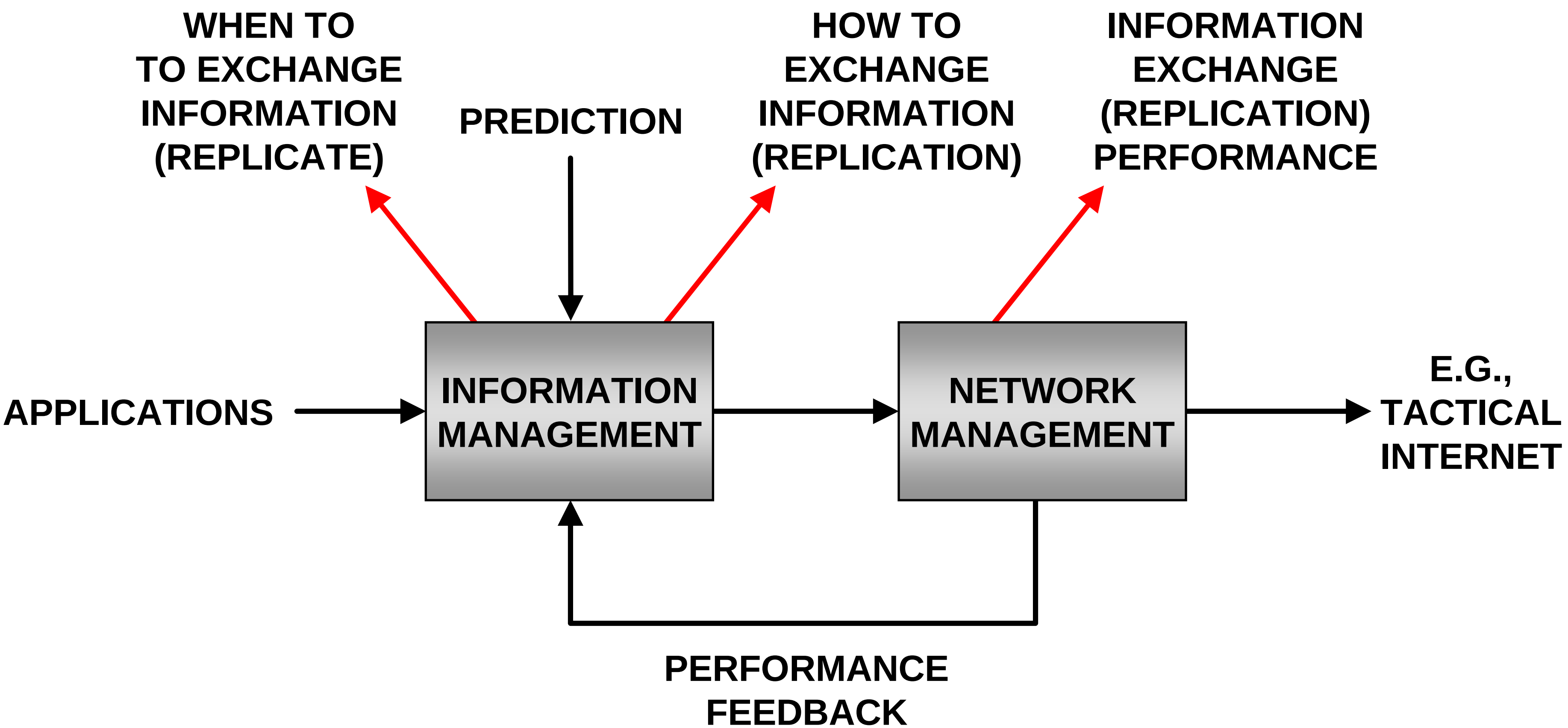


Bandwidth - Prediction Trade-off



Visualizing where one lies on this graph is imperative.

What Can We Do To Accomplish This Goal? Quit Fighting?



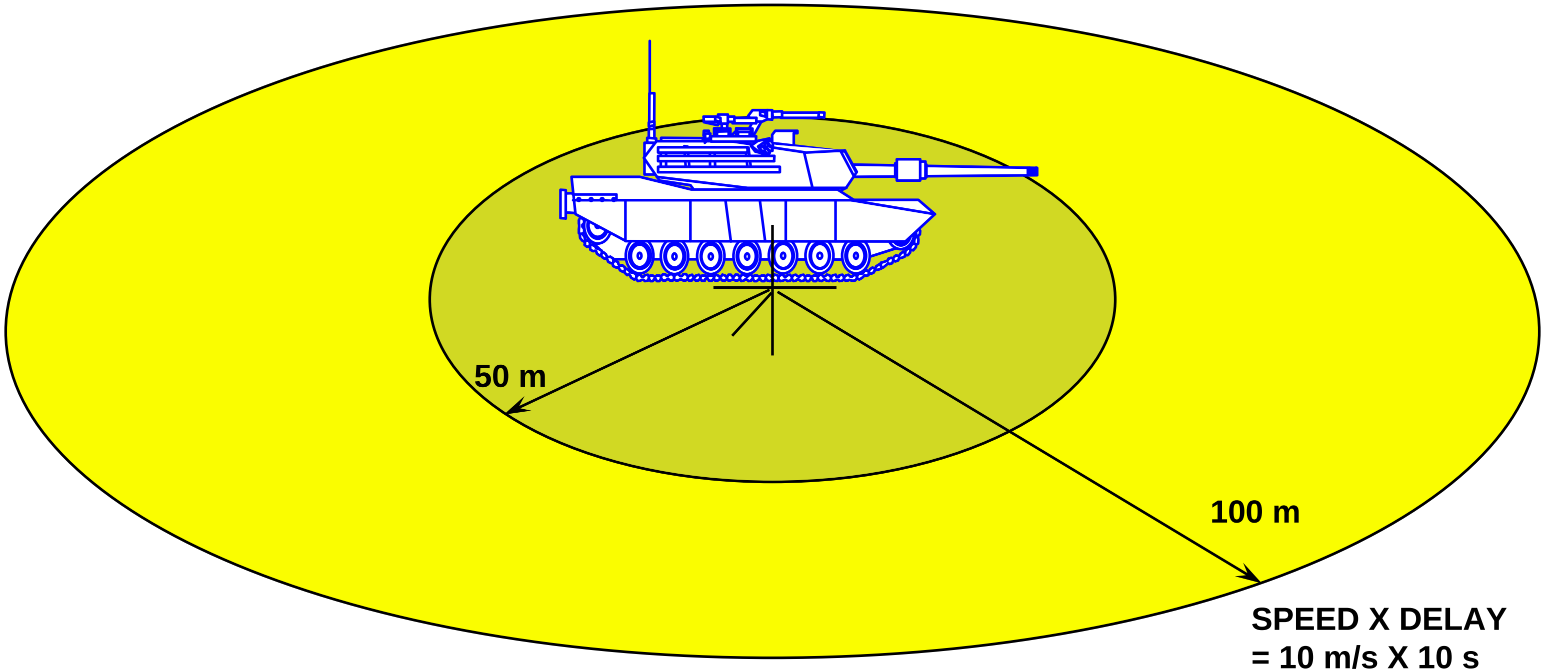
GOAL: BALANCE, OR "TUNE" INFORMATION MANAGEMENT WITH NETWORK RESOURCES

Information Distribution In Tenuous Communications Environments

- **Objective: intelligent & prudent dissemination of information.**
  (to use resources wisely –
  avoid sending any wasteful 1's and 0's).
- **Key Enabler: Model-Based, vs. Message-Based, Battle Command.**
- **Requirements - Perform Information Dissemination that is:**
  Automatic: Hands Off, Context Based.
  Adaptive: Depends on Resources (e.g., Bandwidth or Power).
- **Technologies and Capabilities to Accomplish this Goal.**
  - **Active Database Triggers (now part of most software agents).**
  - **Promiscuous Replication (initiated by triggers).**
  - **Innovative Transport Protocols (e.g., performance feedback).**
- **This means that communications information, like performance and connectivity data, should be part of the data schema.**

Example Of Bandwidth Driven Information Synchronization

SIMPLE CASE: POSITIONAL AWARENESS



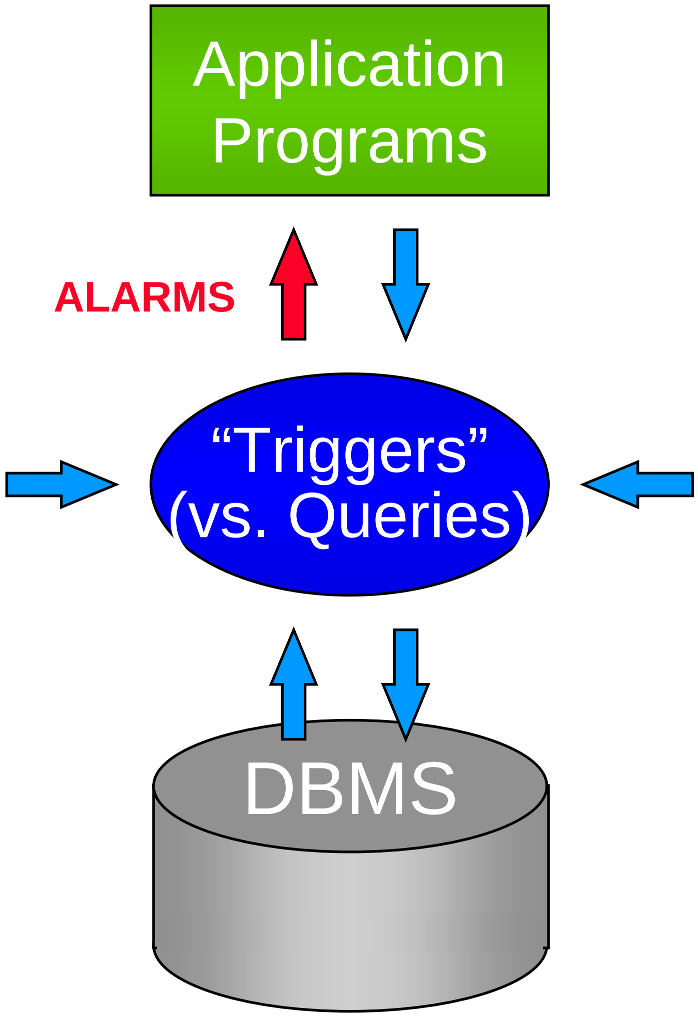
SPEED X DELAY
= 10 m/s X 10 s
= 100 m (best case)

LOCATION

A GROUND SPEED OF 10 METERS / SECONDS (22.5 MPH) WITH AN AVERAGE NETWORK DELAY OF 10 SECONDS MEANS AN UPDATE EVERY 100 METERS AT BEST; OTHERWISE, ONE WILL JUST BACKUP THE OUTPUT QUEUE.

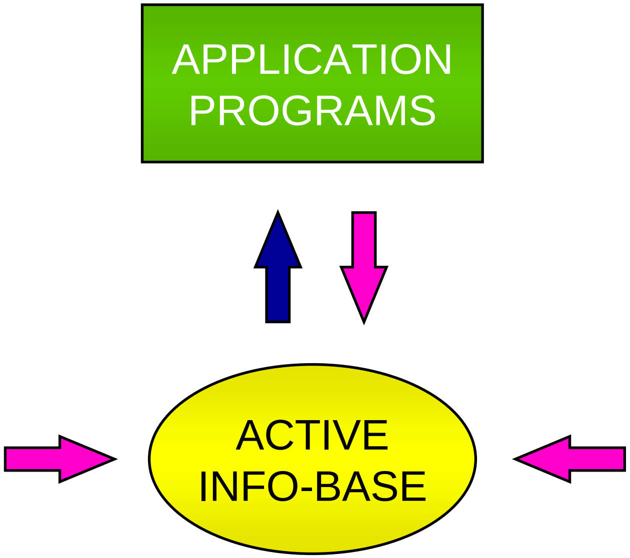
POSITION RESOLUTION VARIES WITH BANDWIDTH CONDITIONS (Vehicle Speed and Network Delay)

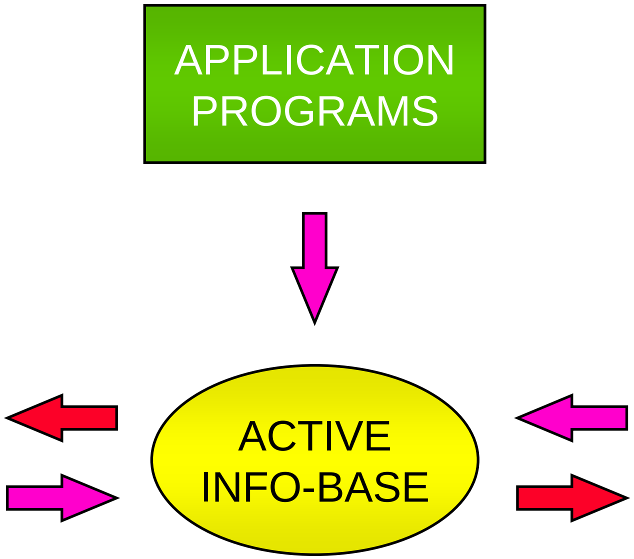
Active Database Concept



In Some Applications, Update Rates Are too Frequent to Reasonably Identify Situations Via Manual Queries. Active Databases Incorporate a Monitor That Allows Predefined Criteria (Triggers) to Be Entered. Incoming Data Are Checked Against the Triggers and Set Off Alarms When Criteria Is Met.

Active Database Triggers Can Invoke Any Action

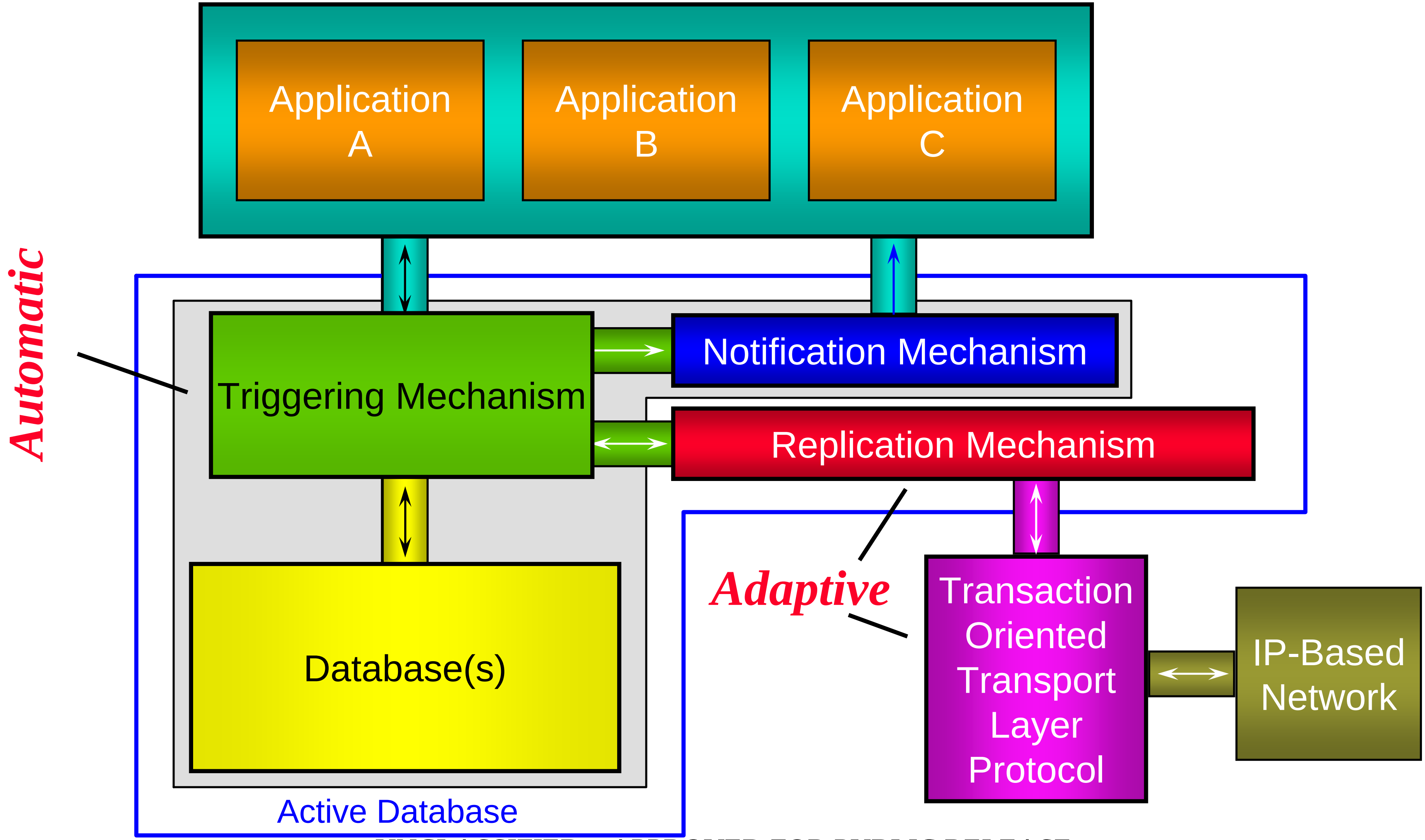




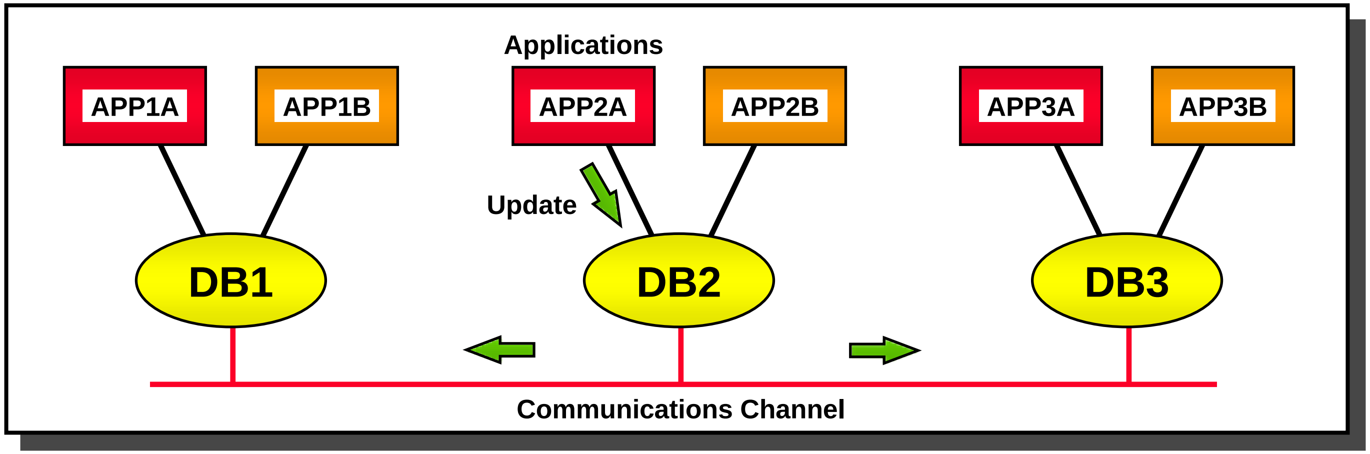
ALARMS
Incoming Information Causes Triggers to Notify Application Programs.

REPLICATION MECHANISM
Incoming Information Causes Triggers to Update Remote Database.

Functional Block Diagram



Some Replication Approaches & Promiscuous Replication



CURRENT REPLICATION TECHNIQUES:

TIGHT CONSISTENCY: ***ALL*** **DATABASES MUST RECEIVE THE UPDATE BEFORE** ***ANY*** **DATABASE CAN APPLY THE UPDATE.**
RESULT: ALL DATABASES PORTRAY IDENTICAL STATES AND HAVE "IDENTICAL" AUDIT TRAILS.

LOOSE CONSISTENCY: LOCAL DATABASE CAN APPLY THE UPDATE BEFORE OTHER DATABASES RECEIVE THE UPDATE;
UPDATES ARE QUEUED AND EVENTUALLY PASSED IN ORDER;
RESULT: ASYNCHRONOUS AUDIT TRAILS.

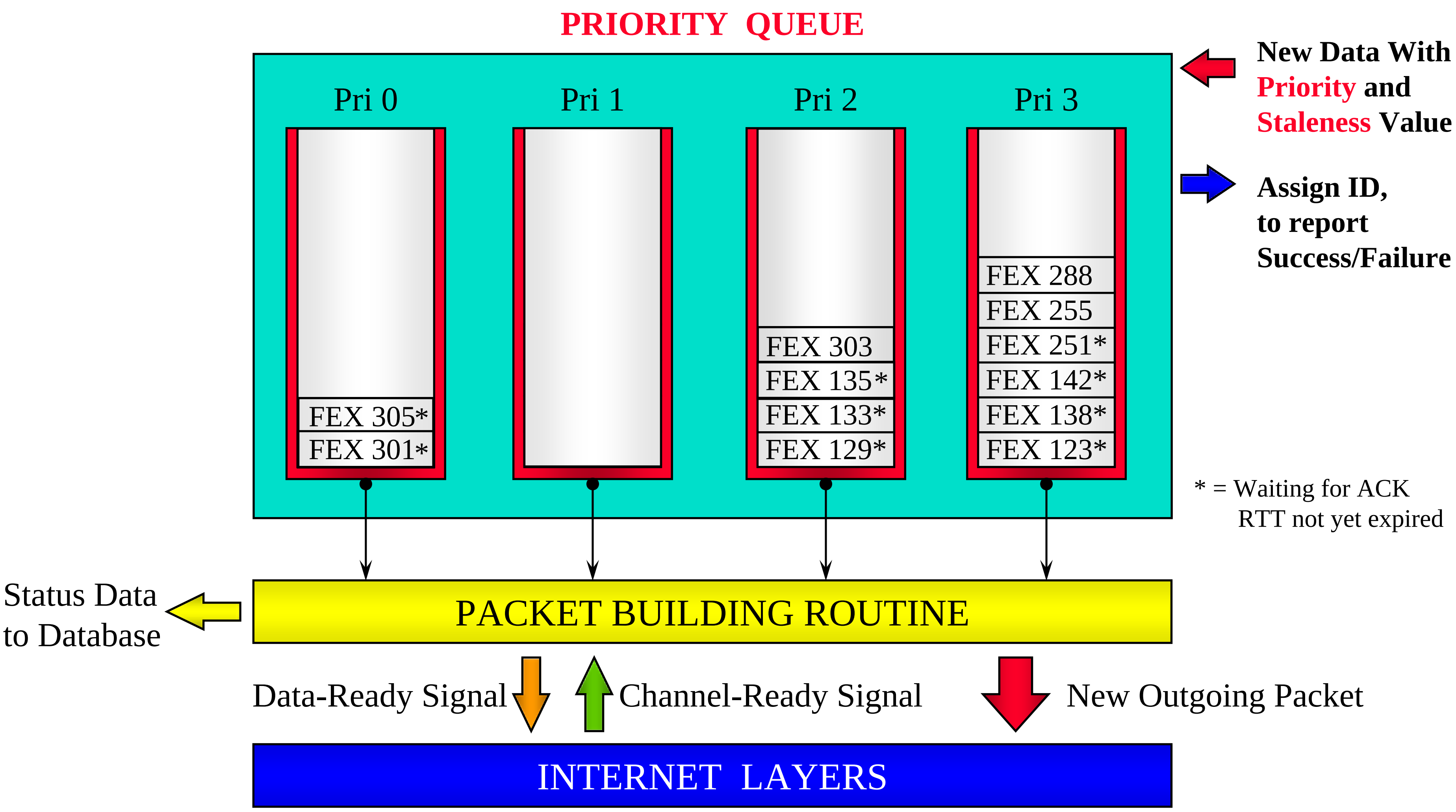
PROMISCUOUS (CASUAL) REPLICATION:
LOCAL DATABASE CAN APPLY THE UPDATE IMMEDIATELY.
REPLICATION IS A FUNCTION OF THE DATABASE STATE AND META-INFORMATION.
(e.g., TACTICAL SIGNIFICANCE AND CURRENT CHANNEL PERFORMANCE.)
COMMON AUDIT TRAILS ARE NOT REQUIRED. SYSTEM PROVIDES BEST EFFORT.

Innovative Protocol Mechanisms

- **Stack Cognizance - Sharing information between ISO layers**
  - **E.g., Signal exchange between Transport & Datalink layers.**
  - **Performance feedback to applications (throughput, delay).**
  - **MTU size to applications to allow context blocking within MTU.**
- **"Just-in-time" Packet Construction**
  - **Context Blocking – packing several small Application PDUs into an MTU.**
  - ***Staleness*: property of data – duration of usefulness.**
  - **Assumes that access delay may be greater than staleness value.**
  - **Transport Layer builds and passes segments when signaled by lower layer (e.g., datalink layer for CSMA).**
- **Overhearing**
  - **Network -> Transport Layer pass-through of packets not addresses to host.**
  - **Overheard packet marked accordingly.**
- **Potential application to Stream Control Transport Protocol (RFC 2690).**

Example: "Just-in-time" Packet Construction

FEX: Fact Exchange – A self-contained, meaningful, atomic Application PDU

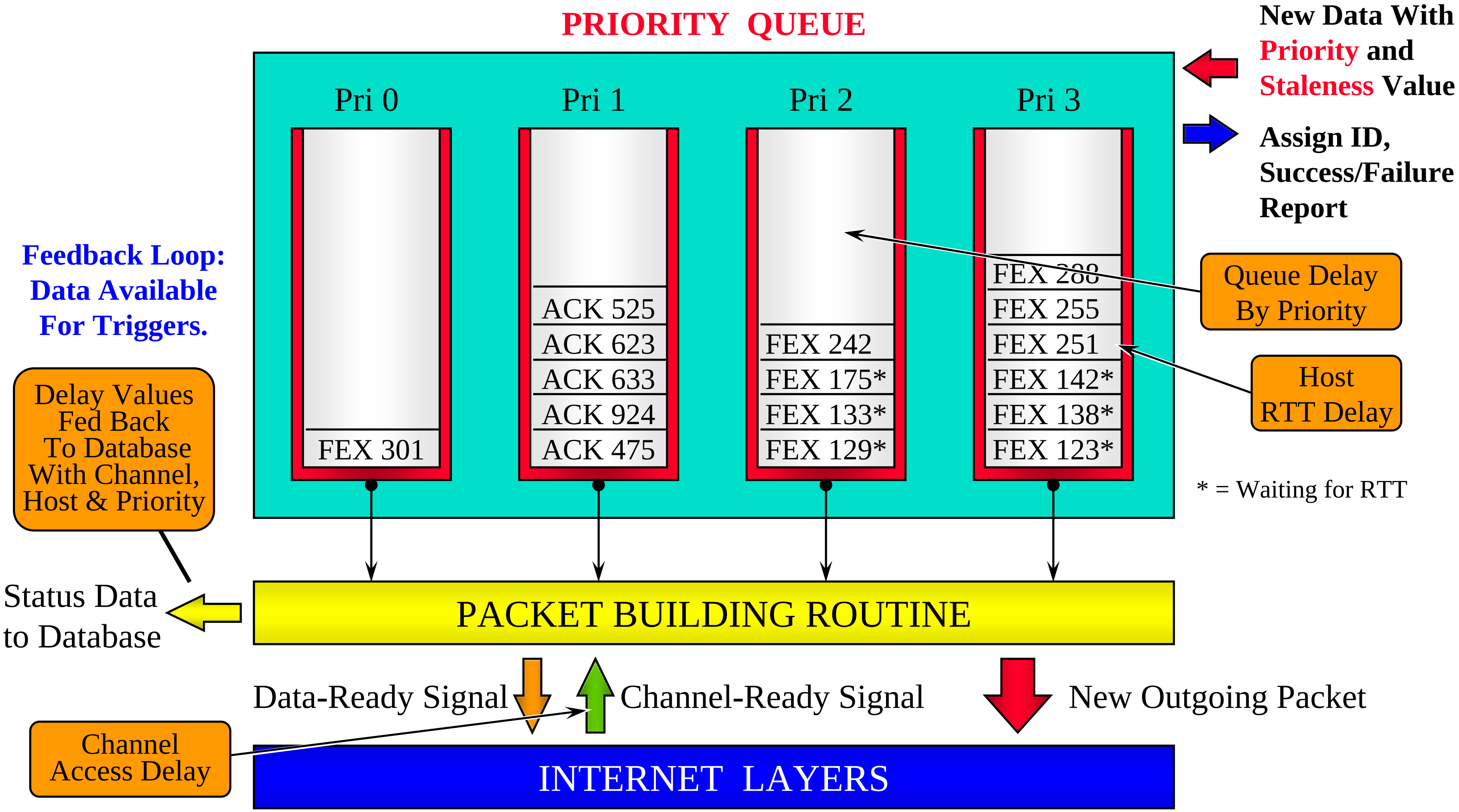


UNCLASSIFIED – APPROVED FOR PUBLIC RELEASE

Resilient Data Transport Features - Interface with Network Management

- Expects that there are more data to send than can be sent; Goals:
  - Never Send An Unnecessary 1 or 0.
  - Always Send Most Important Information First.
  - Expect Delays and Failures.
  - Report Delays and Failures.
- Interface Data Unit Includes
  Priority and *Staleness* Values for Local Use.
- Datagram-Oriented.
  - Multi-destination (reliable) and Broadcast (unreliable) Supported.
  - Concatenation Expected by Lower Layers (e.g., 188-220A).
  - Overhearing.
- *Just-in-Time* Packet Construction (Postpone Selection).
- Performance Measurements Maintained and Reported.

Adaptation - Collection Of Channel Performance Data (Delay Metrics)



Collection of Channel Performance Data
(Various Delay Metrics)

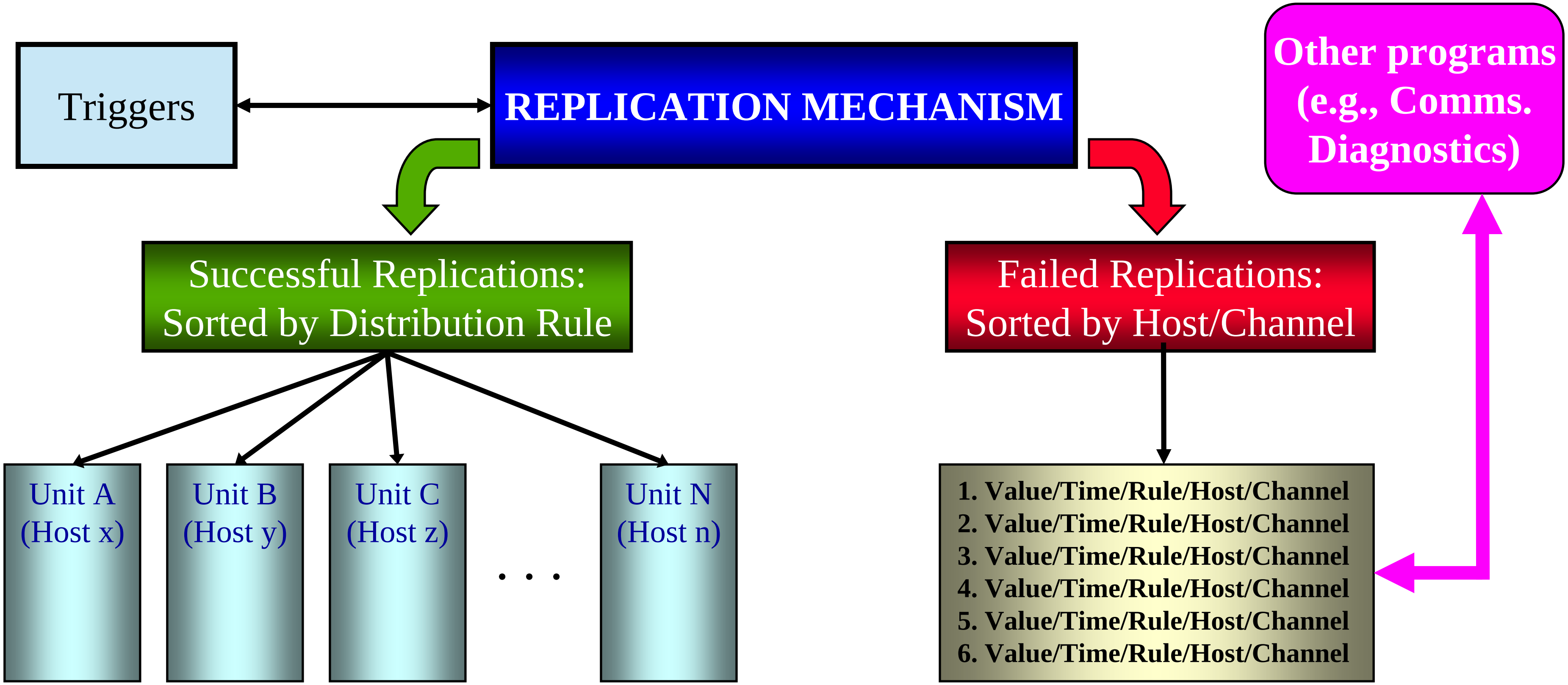
- **Channel delay data collected and reported back to database for use in trigger mechanism (just another parameter):**
  - **Average *Channel Access Delay* for Each Channel Transmission.**
  - **Average *Round Trip Time* for Each Host.**
  - **Average *Queue Delay* for Each Exchange.**
- **Use delay values and failure reports (reasons) to:**
  - **Compute Optimal Reporting Frequencies (see example).**
  - **Tune Information Exchanges to Balance Channel Loading.**
  - **Check for Stable Queue Sizes and Adjust Dynamically.**
- ***Passively* attempt to identify and predict communication connection and congestion problems.**
  - **Look For Patterns of Concern - by Channel and by Host.**
  - **Include A-priori Information (e.g., Battle Plan)**

Realistic Handling Of Failed Replications

Concept: Track & Handle Replication *Successes* Separately from Replication *Failures*.

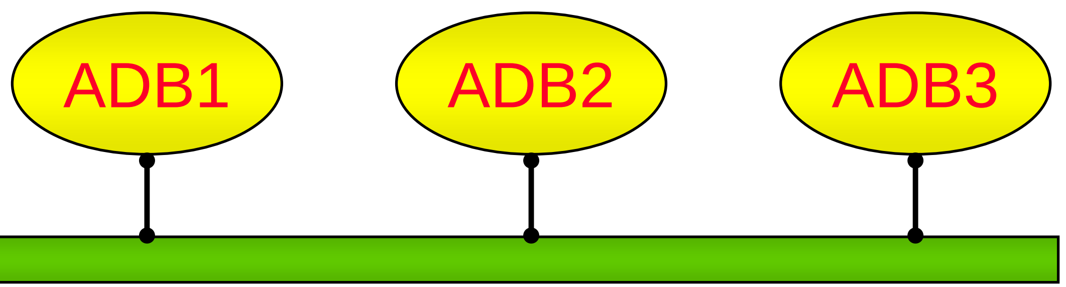
Successes allow one to know what others know about you.

Failures used to derive state of communications system and potential problems.



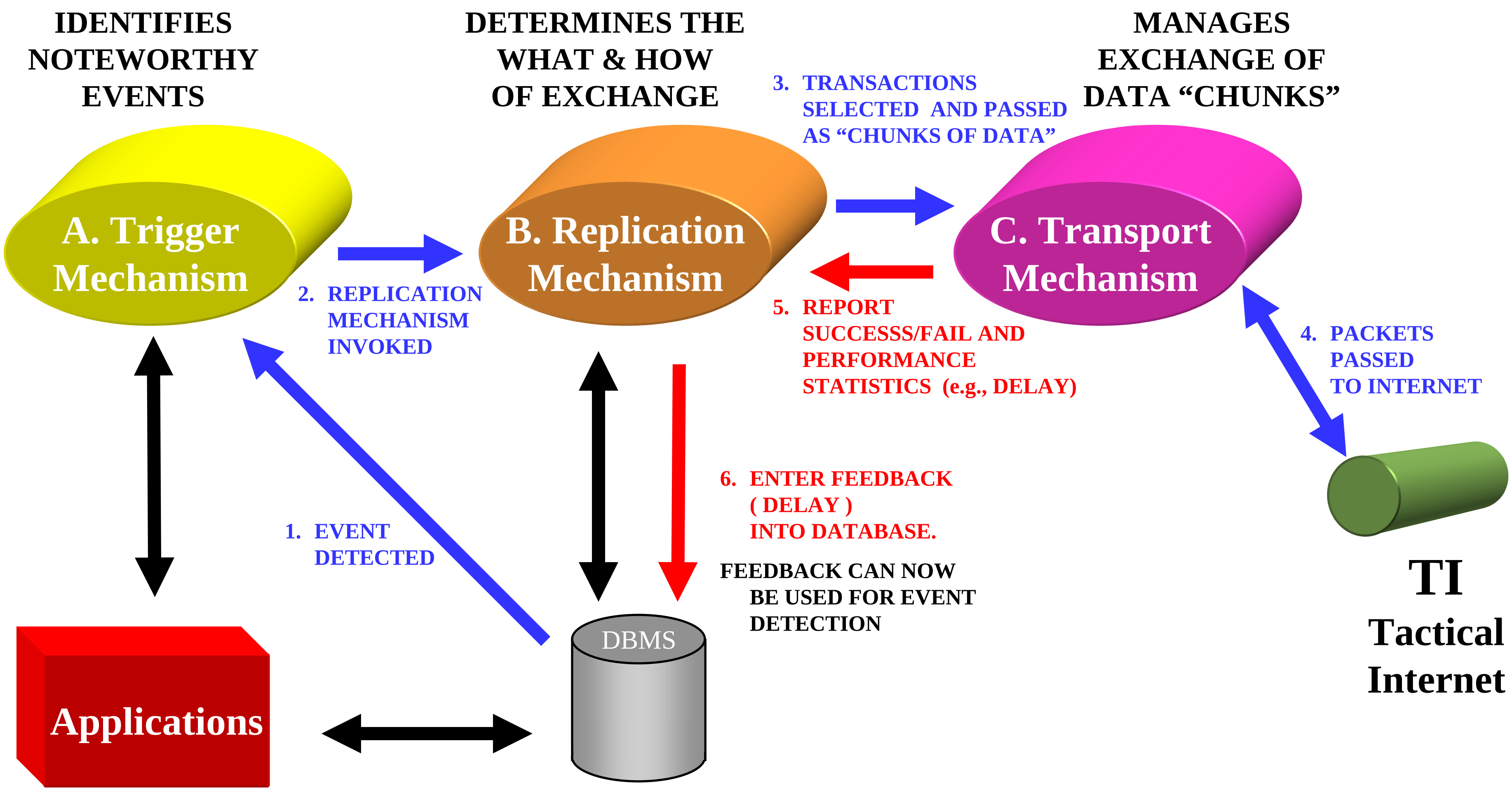
Communication-Based Transaction Failures

* Staleness Timer Expires && Number of Transmissions = 0	Congestion
* Staleness Timer Expires && $0 \leq$ Number of Transmissions $\leq$ Max	Congestion or Channel/Host Failure
* Number of Transmissions > Max	Channel/Host Failure



Communications Channel

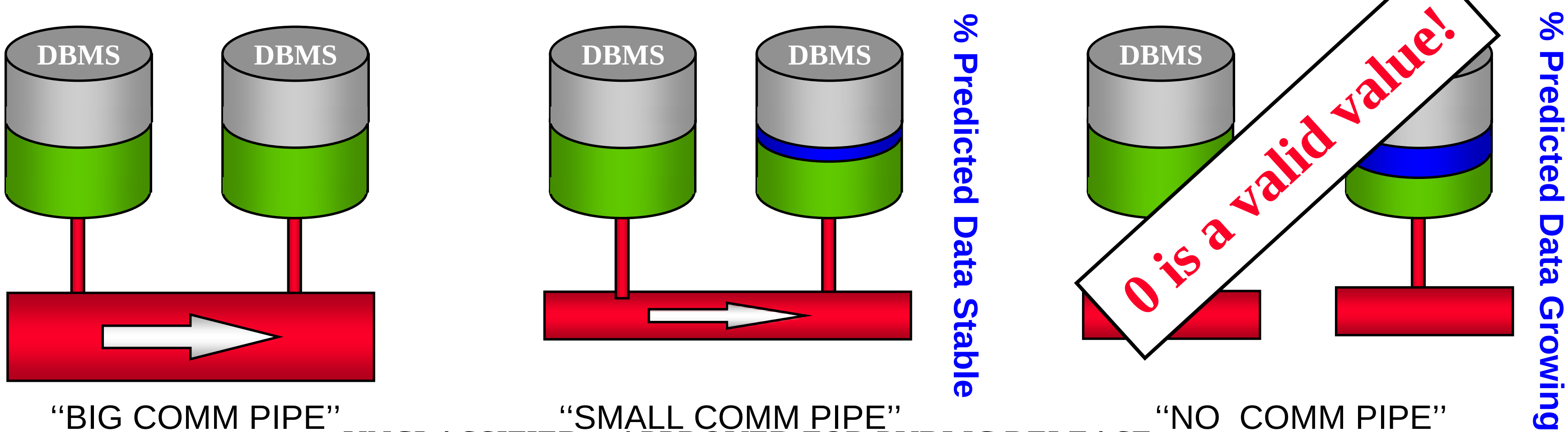
Summary: Relationship Between the Three Processes Using Model-Base Battle Command



Summary of Resilient Information Management & Distribution Features

1. Information Synchronization Can Automatically Adjust to Bandwidth Variations by:
   A. Including Communications in the Data Model; and
   B. Passively Collecting Network Performance Statistics.
2. The Active Database Triggers Can Now Refer to the Statistics to Dynamically Vary Database Synchronization (e.g., Position Report Resolution Vary With Network Delay).
3. Audit Trail Requirements Are Relaxed Between Databases to Ensure that the Most Important Information Gets Distributed (First) – I.e., *Promiscuous* Data Replication.
4. Because a Battlefield Model Exists at Each Node, Sophisticated Application Programs Can Be Used to "Fill Holes" Via Prediction.

Concurrent Databases	Controlled Asynchronization	Unsynchronized Databases
No Prediction Required	Some Prediction Required	Updates Via Prediction



For More Information

Dr. Sam Chamberlain
Computational & Information Sciences Directorate
U.S. Army Research Laboratory (ARL)

Director, USARL
ATTN: AMSRL-CI-CT (Chamberlain)
APG, MD 21005-5067

Phone: 410-278-8948 (DSN 298); Fax: 2934;
Email: wildman@arl.army.mil
URL: http://www.arl.army.mil/~wildman
(Papers Online)